Notes de cours

24 janvier 2023

1 Inertie : changement de point de repère

On cherche à comparer $I_{\mathbf{g}}$ et $I_{\mathbf{a}}$, pour un vecteur $\mathbf{a}$ quelconque.

$$\begin{aligned}
I_{\mathbf{a}} &= \sum_{i=1}^{n} p_i \|\mathbf{e}_i - \mathbf{a}\|_{\mathbf{M}}^2 = \sum_{i=1}^{n} p_i \|\mathbf{e}_i - \mathbf{g} + \mathbf{g} - \mathbf{a}\|_{\mathbf{M}}^2 \\
&= \sum_{i=1}^{n} p_i \|\mathbf{e}_i - \mathbf{g}\|_{\mathbf{M}}^2 + \sum_{i=1}^{n} p_i \|\mathbf{g} - \mathbf{a}\|_{\mathbf{M}}^2 + 2 \times \left\langle \sum_{i=1}^{n} p_i(\mathbf{e}_i - \mathbf{g}), \mathbf{g} - \mathbf{a} \right\rangle_{\mathbf{M}} \\
&= I_{\mathbf{g}} + \|\mathbf{g} - \mathbf{a}\|_{\mathbf{M}}^2 + 2 \times \langle 0, \mathbf{g} - \mathbf{a} \rangle_{\mathbf{M}} = I_{\mathbf{g}} + \|\mathbf{g} - \mathbf{a}\|_{\mathbf{M}}^2.
\end{aligned}$$

On voit donc que $I_{\mathbf{a}}$ n'est différent de $I_{\mathbf{g}}$ que par une valeur fixe. Il ne présente pas d'intérêt.

2 Seconde forme de l'inertie

$$\begin{aligned}
\sum_{i=1}^{n} \sum_{j=1}^{n} p_i p_j \|\mathbf{e}_i - \mathbf{e}_j\|_{\mathbf{M}}^2 &= \sum_{i=1}^{n} \sum_{j=1}^{n} p_i p_j \left[\|\mathbf{e}_i - \mathbf{g}\|_{\mathbf{M}}^2 + \|\mathbf{g} - \mathbf{e}_j\|_{\mathbf{M}}^2 - 2\langle \mathbf{e}_i - \mathbf{g}, \mathbf{e}_j - \mathbf{g} \rangle_{\mathbf{M}} \right] \\
&= 2 \sum_{i=1}^{n} p_i \|\mathbf{e}_i - \mathbf{g}\|_{\mathbf{M}}^2 - 2 \langle \sum_{i=1}^{n} p_i(\mathbf{e}_i - \mathbf{g}), \sum_{j=1}^{n} p_j(\mathbf{e}_j - \mathbf{g}) \rangle_{\mathbf{M}} \\
&= 2 I_{\mathbf{g}} - 2 \times \langle 0, 0 \rangle_{\mathbf{M}}.
\end{aligned}$$

3 Inertie et trace

$$I_{\mathbf{g}} = \sum_{i=1}^{n} p_i \|\mathbf{e}_i - \mathbf{g}\|_{\mathbf{M}}^2 = \sum_{i=1}^{n} p_i (\mathbf{e}_i - \mathbf{g})' \mathbf{M} (\mathbf{e}_i - \mathbf{g}).$$

D'autre part, comme $I_\mathbf{g}$ est un nombre,

$$\begin{aligned}
I_\mathbf{g} = \mathrm{Tr}(I_\mathbf{g}) &= \sum_{i=1}^n p_i \mathrm{Tr}\Big[(\mathbf{e}_i - \mathbf{g})'\mathbf{M}(\mathbf{e}_i - \mathbf{g})\Big] = \sum_{i=1}^n p_i \mathrm{Tr}\Big[(\mathbf{e}_i - \mathbf{g})(\mathbf{e}_i - \mathbf{g})'\mathbf{M}\Big] \\
&= \mathrm{Tr}\Big\{\Big[\sum_{i=1}^n p_i(\mathbf{e}_i - \mathbf{g})(\mathbf{e}_i - \mathbf{g})'\Big]\mathbf{M}\Big\} = \mathrm{Tr}(\mathbf{VM}).
\end{aligned}$$

4 Propriétés spectrales de VM

Soit $\mathbf{T} = \mathrm{diag}(\sqrt{mj})$ la matrice diagonale qui vérifie $\mathbf{T}^2 = \mathbf{M}$. $\mathbf{TVT}$ est une matrice symétrique et possède donc des valeurs propres réelles que l'on classe arbitrairement par ordre décroissant

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_p,$$

et p vecteurs propres réels orthonormés $\mathbf{v}_1, \ldots, \mathbf{v}_p$ associés. De plus, $\mathbf{TVT}$ est une matrice semi-définie puisque, pour tout vecteur $\mathbf{v}$

$$\begin{aligned}
\mathbf{v}'\mathbf{TVTv} &= \sum_{j=1}^p \sum_{\ell=1}^p v_j v_\ell \sqrt{m_j}\sqrt{m_\ell}\,\mathrm{cov}(\mathbf{x}^j, \mathbf{x}^\ell). \\
&= \mathrm{cov}\left(\sum_{j=1}^p v_j\sqrt{m_j}\mathbf{x}^j, \sum_{\ell=1}^p v_\ell\sqrt{m_\ell}\mathbf{x}^\ell\right) = \mathrm{var}\left(\sum_{j=1}^p v_j\sqrt{m_j}\mathbf{x}^j\right) \geq 0
\end{aligned}$$

et ses valeurs propres sont donc positives ou nulles. Si on pose $\mathbf{a}_k = \mathbf{T}^{-1}\mathbf{v}_k$, on a alors

$$\begin{aligned}
\mathbf{TVTv}_k &= \lambda_k \mathbf{v}_k, \\
\mathbf{VTv}_k &= \lambda_k \mathbf{T}^{-1}\mathbf{v}_k, \\
\mathbf{VMT}^{-1}\mathbf{v}_k &= \lambda_k \mathbf{T}^{-1}\mathbf{v}_k, \\
\mathbf{VMa}_k &= \lambda_k \mathbf{a}_k.
\end{aligned}$$

Les $\mathbf{a}_k$ sont $\mathbf{M}$-orthonormés car

$$\langle \mathbf{a}_k, \mathbf{a}_\ell \rangle_\mathbf{M} = \mathbf{a}_k'\mathbf{M}\mathbf{a}_\ell = \mathbf{v}_k'\mathbf{T}^{-1}\mathbf{M}\mathbf{T}^{-1}\mathbf{v}_\ell = \mathbf{v}_k'\mathbf{v}_\ell = \langle \mathbf{v}_k, \mathbf{v}_\ell \rangle.$$

5 Preuve partielle du théorème sur l'ACP

On va se contenter de regarder le résultat en dimension 1, c'est-à-dire en projetant sur un axe.

Expression de la projection Tout d'abord, on cherche la définition de la projection $\mathbf{M}$-orthogonale sur l'axe définit par le vecteur $\mathbf{a}$. L'opérateur $\mathbf{P}$ de projection satisfait pour tout vecteur $\mathbf{v}$

$$\mathbf{Pv} = \mu\mathbf{a}, \qquad\qquad (\mathbf{v} - \mathbf{Pv})'\mathbf{Ma} = 0.$$

On en déduit que $\mathbf{v}'\mathbf{Ma} - \mu\mathbf{a}'\mathbf{Ma} = 0$ et donc $\mu = \mathbf{a}'\mathbf{Mv}/\mathbf{a}'\mathbf{Ma}$. Finalement,

$$\mathbf{P} = \frac{\mathbf{aa}'\mathbf{M}}{\mathbf{a}'\mathbf{Ma}}$$

On vérifie que $\mathbf{P}^2 = \mathbf{P}$ et $\mathbf{P}'\mathbf{M} = \mathbf{MP}'$, ce qui est vrai pour toutes les projections $\mathbf{M}$-orthogonales.

L'inertie du nuage projeté En reprenant le calcul « Inertie et trace », on trouve pour le nuage projeté $\mathbf{XP}$

$$\begin{aligned} I_{\mathbf{g}}(\mathbf{XP}) &= \sum_{i=1}^{n} p_i \|\mathbf{P}(\mathbf{e}_i - \mathbf{g})\|^2_{\mathbf{M}} = \sum_{i=1}^{n} p_i (\mathbf{e}_i - \mathbf{g})'\mathbf{P}'\mathbf{MP}(\mathbf{e}_i - \mathbf{g}) \\ &= \sum_{i=1}^{n} p_i (\mathbf{e}_i - \mathbf{g})'\mathbf{MP}(\mathbf{e}_i - \mathbf{g}) = \cdots = \mathrm{Tr}(\mathbf{VMP}) \\ &= \frac{\mathrm{Tr}(\mathbf{VMaa}'\mathbf{M})}{\mathbf{a}'\mathbf{Ma}} = \frac{\mathrm{Tr}(\mathbf{a}'\mathbf{MVMa})}{\mathbf{a}'\mathbf{Ma}} = \frac{\mathbf{a}'\mathbf{MVMa}}{\mathbf{a}'\mathbf{Ma}}, \end{aligned}$$

où on a utilisé la relation $\mathbf{P}'\mathbf{MP} = \mathbf{MPP} = \mathbf{MP}$ à la seconde ligne.

Maximisation de l'inertie On cherche le $\mathbf{a}$ qui maximise l'inertie. Pour cela on doit dériver $I_{\mathbf{g}}(\mathbf{XP})$ par rapport à $\mathbf{a}$, dans le sens

$$\frac{\partial I_{\mathbf{g}}(\mathbf{XP})}{\partial \mathbf{a}} = \begin{pmatrix} \frac{\partial I_{\mathbf{g}}(\mathbf{XP})}{\partial a_1} \\ \vdots \\ \frac{\partial I_{\mathbf{g}}(\mathbf{XP})}{\partial a_p} \end{pmatrix}$$

Il est facile de se persuader que pour toute matrice $\mathbf{G}$,

$$\frac{\partial \mathbf{a}'\mathbf{Ga}}{\partial \mathbf{a}} = 2\mathbf{Ga},$$

et donc

$$\frac{\partial I_{\mathbf{g}}(\mathbf{XP})}{\partial \mathbf{a}} = \frac{\mathbf{a}'\mathbf{Ma} \cdot 2\mathbf{MVMa} - \mathbf{a}'\mathbf{MVMa} \cdot 2\mathbf{Ma}}{(\mathbf{a}'\mathbf{Ma})^2} = 0,$$

ce qui implique

$$\mathbf{MVMa} = \frac{\mathbf{a}'\mathbf{MVMa}}{\mathbf{a}'\mathbf{Ma}}\mathbf{Ma},$$

et comme $\mathbf{M}$ est inversible, $\mathbf{a}$ est vecteur propre de $\mathbf{VM}$ associé à la valeur propre $\lambda = I_{\mathbf{g}}(\mathbf{XP})$:

$$\mathbf{VMa} = \lambda\mathbf{a}.$$

Finalement, l'axe qui maximise l'inertie projetée est le vecteur $\mathbf{a}_1$ défini en section 4.

6 Variables liées et valeurs propres

On cherche à montrer qu'à chaque fois que des variables sont liées par une relation linéaire, l'analyse en composante principale des données correspondantes produit un vecteur $\mathbf{u}$ associé une valeur propre nulle. On considère pour cela une table $\mathbf{X} = (x_i^j)$ de données avec n individus et p variables, ainsi que sa version centrée réduite $\mathbf{Z} = (z_i^j)$, où comme d'habitude $z_i^j = (x_i^j - \bar{x}^j)/\sigma_j$, $\bar{x}^j$ est la moyenne arithmétique de la variable j et σ_j son écart type. Pour simplifier, on supposera que toutes les variables ont le même poids $1/n$. Dans ce cas, on rappelle que la matrice de corrélation de $\mathbf{X}$ s'écrit $\mathbf{R} = \frac{1}{n}\mathbf{Z}'\mathbf{Z}$.

Dans la suite, on suppose qu'il existe des coefficients non tous nuls $v_1, \ldots, v_p$ et v_0 tels que, pour tout i,

$$v_1 x_i^1 + v_2 x_i^2 + \ldots + v_p x_i^p = v_0.$$

Les moyennes arithmétiques des variables vérifient alors

$$\begin{aligned} v_1\bar{x}^1 + v_2\bar{x}^2 + \ldots + v_p\bar{x}^p &= v_1\frac{1}{n}\sum_{i=1}^n x_i^1 + \ldots + v_p\frac{1}{n}\sum_{i=1}^n x_i^p \\ &= \frac{1}{n}\sum_{i=1}^n \left(v_1 x_i^1 + \ldots + v_p x_i^p\right) \\ &= \frac{1}{n}\sum_{i=1}^n v_0 = v_0. \end{aligned}$$

Les variables centrées vérifient pour chaque individu i

$$v_1(x_i^1 - \bar{x}^1) + v_2(\bar{x}^2 - x_i^2) + \ldots + v_p(\bar{x}^p - x_i^p) = v_0 - v_0 = 0,$$

et comme $x_i^j - \bar{x}^j = \sigma_j z_i^j$, pour les variables centrées réduites on obtient

$$v_1 s_1 z_i^1 + v_2 s_2 z_i^2 + \ldots + v_p s_p z_i^p = 0.$$

Un facteur propre non nul $\mathbf{u}$ associé à la valeur propre nulle est tel que $\mathbf{Ru} = \mathbf{0}$, puisque les facteurs propres sont des vecteurs propres de la matrice de corrélation. On considère le vecteur $\mathbf{u} = (v_1\sigma_1, \ldots, v_p\sigma_p)'$. Le vecteur est évidemment non nul et on peut réécrire matriciellement la relation trouvée précédemment comme $\mathbf{Zu} = \mathbf{0}$.

On a donc $\mathbf{Ru} = \frac{1}{n}\mathbf{Z}'\mathbf{Zu} = \frac{1}{n}\mathbf{Z}'\mathbf{0} = \mathbf{0}$, ce qui répond à la question posée.